

Automatically Determining Allowable Combinations of a Class of Flexible Multiword Expressions

Afsaneh Fazly, Ryan North, and Suzanne Stevenson

Department of Computer Science
University of Toronto
Toronto, ON M5S 3H5
Canada
{afsaneh,ryan,suzanne}@cs.toronto.edu

Abstract. We develop statistical measures for assessing the acceptability of a frequent class of multiword expressions. We also use the measures to estimate the degree of productivity of the expressions over semantically related nouns. We show that a linguistically-inspired measure outperforms a standard measure of collocation in its match with human judgments. The measure uses simple extraction techniques over non-marked-up web data.

1 Light Verb Constructions

Recent work in NLP has recognized the challenges posed by the rich variety of multiword expressions (MWEs) (e.g., Sag et al., 2002). One unsolved problem posed by MWEs is how they should be encoded in a computational lexicon. Many MWEs are syntactically flexible; for these it is inappropriate to treat the full expression as a single word. However, fully compositional techniques can lead to overgeneralization, because flexible MWEs are often *semi*-productive: new expressions can only be formed from limited combinations of semantically and syntactically similar component words. In order to achieve accurate lexical acquisition methods, we must determine computational mechanisms for capturing the allowable combinations of such MWEs.

Our focus here is on light verb constructions (LVCs); these are largely compositional and semi-productive MWEs having a high frequency of occurrence across many diverse languages (Karimi, 1997; Miyamoto, 2000; Butt, 2003). LVCs combine a member of a restricted set of light verbs, such as *give*, *take*, and *make* among others in English, with a wide range of complements of varying syntactic categories. We consider a common class of LVCs, in which the complement is a noun generally used with an indefinite article, as in (a–c) below:

- | | |
|--|--|
| a. Priya <i>took a walk</i> along the beach. | d. Priya <i>walked</i> along the beach. |
| b. Allene <i>gave a smile</i> when she saw us. | e. Allene <i>smiled</i> when she saw us. |
| c. Randy <i>made a joke</i> to his friends. | f. Randy <i>joked</i> to his friends. |

Moreover, the complement nouns in these expressions, such as *walk*, *smile*, and *joke* in (a–c), have a stem form identical to a verb. Because the light verb is “semantically bleached” to some degree (Butt, 2003), most of the meaning of these LVCs comes from